

Modelos Probabilistas Aplicados

Johanna Bolaños Zúñiga

Matricula: 1883900

Tarea 8

1. Discusión sobre material proporcionado en clase

El teorema de Bayes permite que el proveedor de atención médica convierta los resultados de una prueba en la probabilidad de tener una enfermedad. Este teorema se utiliza para calcular la probabilidad condicional de un suceso, teniendo información de antemano de dicho suceso, en otras palabras, sirve para determinar la probabilidad de una de las causas, puesto que ya se observó el efecto o suceso, y se calcula de la manera siguiente:

$$P(A_n | B) = \frac{P[B | A_n] * P[A_n]}{\sum_{i=1}^n P[B | A_i] * P[A_i]}, \quad (1)$$

donde B es el suceso sobre el que tenemos información previa, A_n son los distintos sucesos condicionados y A_i es la causa i , donde $i = 1, 2, 3, \dots, n$. Aplicando la ley de la multiplicación a la ecuación 1 se obtiene la siguiente expresión (ecuación 2):

$$P(A | B) = \frac{P[B | A] * P[A]}{P[B]}, \quad (2)$$

donde en la parte del numerador se tiene la probabilidad condicionada, y en la parte del denominador la probabilidad total.

De acuerdo a la enfermedad del Covid-19, hay diversos estudios donde se utiliza el teorema de Bayes para analizar los casos que se tienen al realizar la prueba y de los factores que influyen para su cálculo. Los casos que se pueden presentar son:

- Verdaderamente positivos: si una persona con Covid-19 da positivo a este.
- Verdaderamente negativos: cuando una persona sin Covid-19 da negativo en las pruebas.

-
- Falsos positivos: cuando una persona sin Covid-19 da positivo a este.
 - Falsos negativos: si una persona con Covid-19 da negativo en la prueba.

Para interpretar con una mejor precisión el resultado de la prueba, es necesario conocer su valor predictivo positivo (VPP)¹ y negativo (VPN)², los cuales dependen de la sensibilidad y especificidad y la prevalencia o la probabilidad previa a la prueba (probabilidad a priori). Se conoce como sensibilidad de cualquier prueba biomédica a la probabilidad de que una persona dé positivo dado que tiene la enfermedad y, la especificidad a la probabilidad de que una persona dé negativo dado que no tiene la enfermedad. En el trabajo de Schnipper [10], de manera interactiva muestran como calcular los VPP y VPN.

Por ejemplo, en la investigación de Lewis [8], se utiliza el teorema de Bayes para demostrar cómo la calidad de una prueba de Covid-19 depende de la magnitud del brote (tasa base), en este estudio, reescriben la ecuación 2 usando las probabilidades relevantes para las pruebas de Covid-19 dando como resultado la ecuación 3,

$$P(Cov | Pos) = \frac{P(Pos | Cov) * P(Cov)}{P(Pos | Cov) * P(Cov) + P(Pos | NoCov) * P(NoCov)}, \quad (3)$$

donde $P(Cov | Pos)$ es la probabilidad de que una persona tenga Covid-19 dado que ha dado positivo en la prueba, $P(Pos | Cov)$ es la sensibilidad de la prueba, $P(Cov)$ es la tasa base o la prevalencia (número de casos de Covid-19 / número total de la población a analizar), $P(NoCov)$ el porcentaje de personas que no tienen la enfermedad, $(1 - P(Cov))$ y $P(Pos | NoCov)$ es la probabilidad de los falsos positivos.

De acuerdo a la ecuación 3, considero que es correcto determinar que, incluso cuando se utiliza una prueba muy sensible, cuanto menor sea la tasa base de la enfermedad, más probabilidad se tiene de obtener falsos positivos. Asimismo, entre mayor sea la tasa base, más probabilidad hay de obtener falsos negativos a medida que se aumenten las pruebas.

Por otro lado, en el trabajo realizado por Ranjan [9], propone un modelo donde calculan el valor predictivo positivo VPP y la prevalencia o tasa base está determinada por la ecuación 4,

$$P = 1 - (1 - p)^x, \quad (4)$$

donde P es la probabilidad de que al menos 1 persona de las x cantidad de muestras tenga Covid-19 y p es la probabilidad de que 1 persona tenga la enfermedad. En este trabajo, el hecho de que la prevalencia sea igual al número de casos de Covid-19 dividido el número total de la población a analizar, demuestra que el cálculo del VPP, no es un indicador de gran ayuda para determinar el porcentaje de personas realmente están infectadas, aún sí se considera que las personas a evaluar viven en un área donde es más

¹El valor predictivo positivo (VPP) permite identificar el porcentaje de personas que tienen la enfermedad cuando la prueba ha dado positiva.

²El valor predictivo negativo (VPN) de una prueba es la “capacidad de descartar una enfermedad” dado un resultado negativo de la prueba.

probable que contraiga el virus (zona roja), por lo tanto, la metodología propuesta en cuanto a determinar los verdaderos positivos con base en el número de muestras realizadas, no parece erróneo. Sin embargo, considero que el estudio realizado carece de claridad respecto a los resultados obtenidos.

En la investigación de Good [6], aplican un análisis bayesiano (cálculo del VPP y VPN) para ilustrar la interpretación de las pruebas negativas de Covid-19 considerando que la prevalencia es la probabilidad previa a la prueba, es decir, una probabilidad con base en la sospecha clínica de padecer la enfermedad. Proponen dos escenarios (alta y baja probabilidad previa a la prueba de infección por Covid-19) clínicos. Para ambos escenarios, asumen una especificidad del 99,9 % y varían la sensibilidad del 70 % al 90 %. Considero que los resultados obtenidos van acorde a la metodología planteada, puesto que se realizó de manera correcta la interpretación de los VPN y VPP.

Finalmente, en la investigación de Chan [3], se realiza el estudio basado en las razones de verosimilitud negativa (LR-, por sus siglas en inglés de, *negative likelihood ratio*) y de odds (razón de probabilidades), los cuales son posibles de calcular utilizando el teorema de Bayes y, la prevalencia o probabilidad previa de la prueba es una probabilidad a priori, es decir, con base en la experiencia del médico y la información que posee acerca de la enfermedad en particular del paciente. Este estudio fue realizado con el fin de obtener una mejor estimación del riesgo que tiene un determinado paciente en tener o contraer una enfermedad cuando el resultado de la prueba es negativo. Cabe aclarar que el cálculo de los LR dependen del tipo de análisis que se desee realizar [5]. El **LR-**, los **odds** y la conversión de estos en **probabilidad**, se calculan mediante las ecuaciones 5, 6 y 7, respectivamente,

$$LR- = \frac{1 - \text{sensibilidad}}{\text{especificidad}}, \quad (5)$$

$$\text{odds} = \frac{\text{probabilidad previa}}{1 - \text{probabilidad previa}}, \quad (6)$$

$$\text{probabilidad} = \frac{\text{odds}}{\text{odds} + 1}. \quad (7)$$

Sin embargo, en lugar de realizar los cálculos de las ecuaciones 6 y 7, se puede utilizar el nomograma de Fagan, el cual proporciona una estimación visual de las probabilidades posteriores a la prueba con base en los *LR*. En la figura 1 se muestra el nomograma de Fang basado en el estudio realizado por Aznar-Orovala [1].

En conclusión, un resultado negativo o positivo sin considerar los antecedentes clínicos, factores genéticos o datos que se conoce con precisión del paciente puede limitar la capacidad de los médicos para realizar acciones y disposiciones apropiadas. Para todas las pruebas de detección, ya sea para Covid-19 u otros diagnósticos, la comprensión de los valores predictivos y las razones de probabilidad con la ayuda del teorema de Bayes, podrían garantizar una interpretación sólida y las recomendaciones y acciones resultantes por parte de los médicos y las partes interesadas.

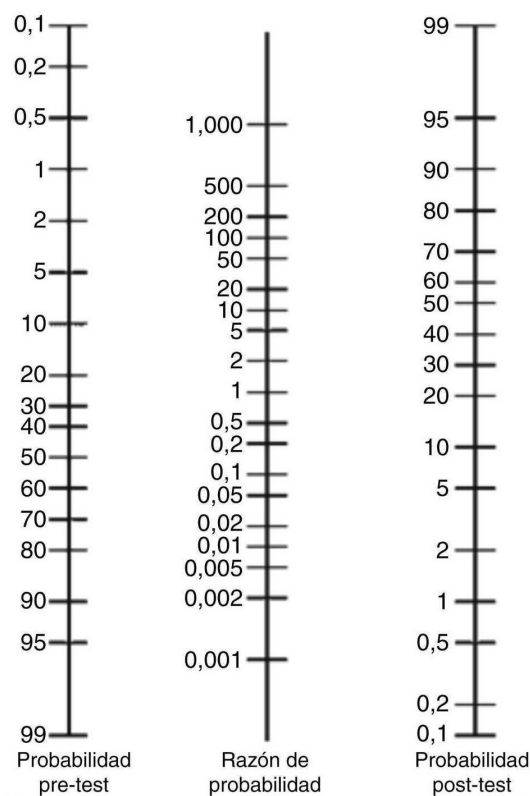


Figura 1: Nomograma de Fagan, donde línea izquierda son las probabilidades previas a la prueba, la línea central es el LR y línea derecha son las probabilidades de tener la enfermedad después del uso de la prueba.

2. Caso práctico

Para la aplicación del teorema de Bayes, en este trabajo se realizó un análisis para estimar de que tan probable es que un paciente con prueba positiva por Covid-19 pueda tener la enfermedad o si el resultado es negativo, que tan probable fue que no padeciera de ella, para lo cual se utilizan los valores predictivos positivos (VPP) o negativos (VPN). Para el cálculo de estos valores, se hace necesario conocer datos como la prevalencia estimada de la enfermedad, la sensibilidad y especificidad de la prueba. Dado que las pruebas por Covid-19 en México se han confirmaron mediante RT-PCR en tiempo real y, aproximadamente, a partir del mes de septiembre el IMMS empezó a utilizar pruebas rápidas en sus hospitales con el objetivo es diferenciar rápidamente el Covid-19 de la influenza, se investigó tanto la sensibilidad como especificidad que tienen ambas pruebas.

De acuerdo a Díaz-Jiménez [4], las pruebas PCR tienen una especificidad cercana al 100 % y una sensibilidad entre el 60 % y 80 % y, para las pruebas rápidas la sensibilidad más baja es del 20 %. Sin embargo, de acuerdo a las especificaciones de las pruebas PCR, se indica que la sensibilidad alcanza un 99 %. Para hallar la prevalencia, según la investigación de Lewis [8], se calcula de acuerdo a la magnitud de

brote y la determina mediante la ecuación 8. Por otro lado, en la investigación de Good [6], la prevalencia podría ser también una probabilidad previa a la prueba, es decir, una probabilidad de la enfermedad con base en la sospecha clínica,

$$\text{prevalencia} = \frac{\text{Total casos por Covid-19}}{\text{Población total}}. \quad (8)$$

Para efectos del presente trabajo, se consideraron dos escenarios con respecto al valor de la prevalencia. En el escenario 1, la prevalencia se calculó mediante la ecuación 8, por lo tanto, de acuerdo a la base de datos proporcionada por Hasell [7] a octubre 25 de 2020, hay 891,160 total de casos por Covid-19 y la población de México es de, aproximadamente, 128,932,753 habitantes, entonces, la prevalencia para el escenario 1 es igual a 0,691 % (0.00691).

Para el escenario 2, se contemplan dos prevalencias o probabilidades previas a la prueba por infección con Covid-19. Una probabilidad de prevalencia baja del 2 % y una alta del 80 %.

Para ambos escenarios se asumió una sensibilidad para las pruebas rápidas del 20 % y para las pruebas PCR la sensibilidad del 80 % y 99 %, La especificidad tanto para los escenarios, como para los tipos de pruebas varían entre el 99 % y 99,9 %. El cálculo los valores predictivos positivos y negativos se realizaron mediante las ecuaciones 9 y 10, respectivamente, basado en la investigación de [5],

$$\text{VPP} = \frac{\text{sensibilidad} * \text{prevalencia}}{\text{sensibilidad} * \text{prevalencia} + (1 - \text{sensibilidad}) * (1 - \text{especificidad})}, \quad (9)$$

$$\text{VPN} = \frac{\text{especificidad} * (1 - \text{prevalencia})}{(1 - \text{sensibilidad}) * \text{prevalencia} + \text{especificidad} * (1 - \text{prevalencia})}. \quad (10)$$

Es importante tener en cuenta que los VPP estiman la probabilidad condicional de que la persona pueda estar enferma cuando la prueba ha dado positiva y los VPN es la “capacidad de descartar una enfermedad” dado un resultado negativo en la prueba, por lo tanto, para determinar la probabilidad condicional de que una persona pueda estar enferma dado que su prueba es negativa (falso negativo) será igual a $1 - \text{VPN}$.

Considerando lo antes mencionado y los resultados del cuadro 1, observamos que al comparar la efectividad o sensibilidad del 99 % en un resultado positivo a través de las pruebas de PCR, con la probabilidad condicional, si se considera la prevalencia con respecto a la magnitud del brote (prevalencia muy baja (0.691 %), da como resultado una estimación de la efectividad de la prueba del 41 % y 87 %, con especificidades del 99 % y 99.9 %, respectivamente. Mientras que, si se considera una prevalencia alta, es decir, con base al riesgo al que ha estado expuesta la persona, la probabilidad condicional estimada para la efectividad de la prueba oscila entre el 96 % y 100 %. Por lo anterior, se podría llegar a un resultado contraintuitivo, ya que la probabilidad de que una persona con un resultado positivo a través de las pruebas PCR se pueda determinar como un posible caso verdadero con Covid-19, oscila entre un 41 % y 87 % y no con una probabilidad del 99 % como se ha especificado para este tipo de pruebas, aunque este resultado depende de la prevalencia considerada, como se observa en la figura 2.

Cuadro 1: Estimación de la probabilidad condicional para las pruebas por Covid-19

Escenario	Prevalencia (%)	Especificidad (%)	Sensibilidad (%)	Teorema de Bayes	
				Prueba positiva (%)	Prueba negativa (%)
1	0.691	99.0	20	12.2	0.6
			80	35.8	0.1
			99	40.8	0.0
		99.9	20	58.2	0.6
			80	84.8	0.1
			99	87.3	0.0
2	2	99.0	20	3.9	1.8
			80	14.0	0.5
			99	16.8	0.0
		99.9	20	80.3	1.6
			80	94.2	0.4
			99	95.3	0.0
	80	99.0	20	98.8	76.4
			80	99.7	44.7
			99	99.7	3.9
		99.9	20	99.9	76.2
			80	100.0	44.5
			99	96.2	3.8

De igual manera, cuando los resultados de las pruebas son negativos, se observa que mientras la prevalencia aumenta, también aumenta la probabilidad de falsos negativos, siendo, principalmente, un foco de atención las pruebas con sensibilidad muy baja, ya que se muestra una probabilidad muy alta (aproximadamente del 76 %) de que las personas con resultados negativos por Covid-19, no estén posiblemente sin la enfermedad, es decir, un aumento de la tasa en los falsos negativos. Por lo tanto, un resultado negativo obtenido por las pruebas rápidas (sensibilidad del 20 %), no sería un buen indicador para descartar la existencia de la enfermedad, ya que como podemos observar en el cuadro 1, con una prevalencia baja o alta, se presenta la probabilidad más alta de falsos negativos. No obstante, considerando que el resultado sea positivo, es posible que puedan ser de utilidad ya que la probabilidad condicional estimada es relativamente buena comparada las pruebas PCR a medida que aumenta la prevalencia, como se observa en la figura 2.

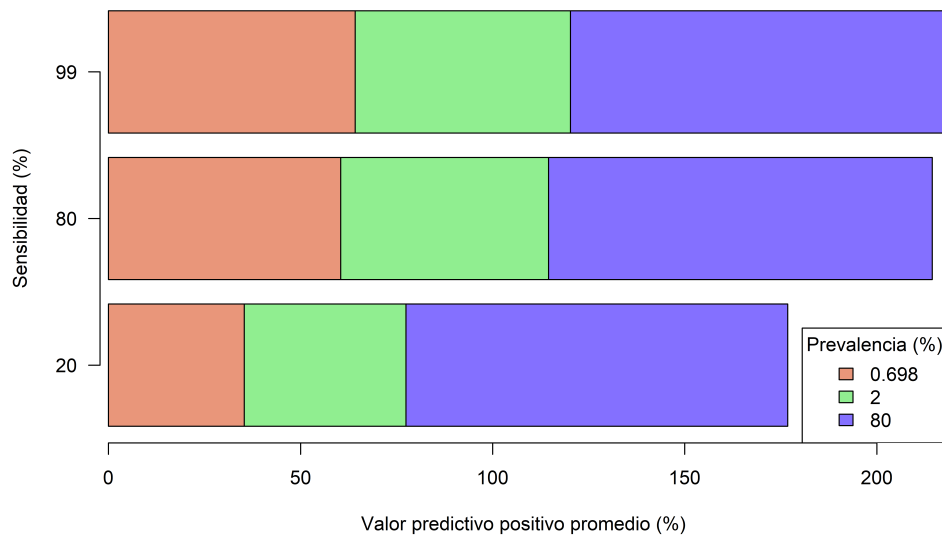


Figura 2: Valores predictivos positivos promedio

Como conclusión general, para las pruebas PCR, mediante el teorema de Bayes, es decir, por medio de los VPP y VPN si se considera un muestreo masivo de la población, determinar si la eficiencia de la prueba es alta, depende de la prevalencia real de la enfermedad del lugar donde se aplica, en este caso, de la magnitud del brote. Sin embargo, cuando el muestreo se considera con base a sectores donde se ha comprobado una mayor prevalencia o riesgo de contagio, la eficiencia (sensibilidad) reportada de la prueba ayuda a estimar con una mayor probabilidad la detección de casos en la población y, de esta manera aplicar las medidas de control necesarias. La base de datos y el código en R utilizado, se encuentran disponibles en el repositorio de GitHub [2].

Referencias

- [1] Aznar-Orovala, E., Mancheño-Alvarob, A., García-Lozanoa, T., Sánchez-Yepes, M. Likelihood ratio and Fagan's nomogram: 2 basic tools for the rational use of clinical laboratory tests. *Revista de Calidad Asistencial*, 28(6):390–391, 2013.
- [2] Bolaños Z., Johanna. Repositorio en GitHub de la clase de modelos probabilistas aplicados. Recursos libre, disponible en github.com/JohannaBZ/Probabilidad/tree/master/Tarea8, 2020.
- [3] Chan, Gar Ming. Interpreting COVID-19 Test Results: a Bayesian Approach. *Interpreting COVID-19 Test Results: a Bayesian Approach*, 06 2020.
- [4] Díaz-Jiménez, Irma Virginia. Interpretation of diagnostic tests for the SARS-Cov-2 virus. *Acta Pediatr Mex*, 41(Supl 1):S51–S57, 2020.
- [5] Fernández R., Raúl. Bayes's theorem and its use in diagnostic test lectures in clinical laboratory. *Revista Cubana de Investigaciones Biomédicas*, 28(3):158–165, 2009.
- [6] Good, Chester B., Hernandez, Inmaculada, Smith, Kenneth. Interpreting COVID-19 Test Results: a Bayesian Approach. *Journal of General Internal Medicine*, 35(8):2490–2491, 06 2020.
- [7] Hasell, J., Mathieu, E., Beltekian, D. A cross-country database of COVID-19 testing. *Sci Data* 7, 345. Recurso disponible en, <https://ourworldindata.org/coronavirus-testing>, 2020.
- [8] Lewis, Michael A. Bayes' theorem and Covid-19 testing. *Significance*, 04 2020.
- [9] Ranjan, Archit. COVID-19, Bayes' theorem and taking probabilistic decisions. Recurso disponible en, <https://towardsdatascience.com/covid-19-bayes-theorem-and-taking-data-driven-decisions-part-1-b61e2c2b3bea>, 2020.
- [10] Schnipper, Effrey L., Sax, Paul E. Covid-19 test accuracy supplement: The math of Bayes' Theorem. Recurso disponible en, <https://www.statnews.com/2020/08/20/covid-19-test-accuracy-supplement-the-math-of-bayes-theorem/>, 2020.